DOCUMENT RESUME

ED 076 631 TM 002 636

AUTHOR Miller, Stuart
TITLE Report on METHODS OF Evaluating Students at the University of California, Berkeley.
PUB DATE Nov 65
NOTE 19p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Academic Achievement; Achievement Rating; College Environment; *Evaluation Methods; *Grades (Scholastic); *Grading; Higher Education; *Pass Fail Grading; *Student Evaluation; Technical Reports

ABSTRACT

Several arguments are made for changing the present system of grading in college. The arguments deal with: (1) the negative effect of grading on the teaching situation--the system is capricious and variable and encourages "grade grubbing"; (2) the lack of accuracy and uniformity of grades; (3) the effect of grading, as an enforcer of the work-success ethic, on the values of the university; (4) the failure of grading to stimulate motivation to learn; (5) the emphasis on grades for graduate admission despite the fact that grades do not predict success in graduate school; (6) the fact that grades are less important in obtaining employment than faculty evaluations; (7) the way in which the grading system encourages a kind of competition alien to the real purposes of university education and discourages the development of intrinsic and lasting intellectual interests and of self-definition and evaluation; (8) the low correlation between grades and creativity; (9) the way the system encourages cheating; (10) the protection grading affords bad teaching. Alternatives to grading suggested are: (1) pass-fail grading--minor enlargement of this option, pass-fail in the breadth requirements, in all nonmajor courses, in the first two years, in all four years; and (2) comprehensive examinations--a system for qualifying examinations for majors and for breadth requirements. (KM)

FORM 6510 PRINTED IN U.S.A.

FILMED FROM BEST AVAILABLE COPY

GF
Miller, Stuart

Jacob B. Michaelsen

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED 076631

REPORT ON METHODS OF EVALUATING STUDENTS AT THE UNIVERSITY OF CALIFORNIA BERKELEY

(This report is part of a longer study prepared for the Select Committee on Education of the Berkeley Division of the Academic Senate by Professor Stuart Miller. It has been edited by Bruce Spiegelberg and condensed by Professor T. Vermeulen, the views and recommendations herein being still Professor Miller's. It is presented in this form for your information and comment.)

November, 1965

TM 002 636

DEC 21 RECD

CONTENTS

I. ARGUMENTS FOR CHANGING THE PRESENT SYSTEM 1

A. Grading and the Teaching Situation 1

B. Grading - Accuracy and Uniformity 3

C. Grading and the Values of the University 3

D. Grading and Motivation 4

E. Grading and Graduate Admission 5

F. Grades and Transfers 7

G. Grades and Employment 7

H. Grades and the Battle of Life 7

I. Grades and Creativity 8

J. Grading and Cheating 9

K. Grading and Teaching 10

II. PROPOSALS 10

A. Pass-Fail Grading 11

1. Minor Enlargement of the Pass-Fail Option 12

2. Pass-Fail in the Breadth Requirements 12

3. Pass-Fail in all Nonmajor Courses 12

4. Pass-Fail in the First Two Years 13

5. Pass-Fail in all Four Years 13

B. Comprehensive Examinations 13

FOOTNOTES 15

I. ARGUMENTS FOR CHANGING THE PRESENT SYSTEM

Professor: What did you get out of the course?

Student: I got a B.

"Marks and terms are clumsy devices, more suitable for measuring cordwood than culture."

David Starr Jordan, The Trend of the American University.

A. Grading and the Teaching Situation

While many people in higher education will defend the present system, few would say it is perfectly satisfactory. The degree of dissatisfaction is greatest among those who are actively concerned with college teaching: those who study education per se, and those teachers who have spent much time thinking about education. Most arguments against the present system of grading done by individual instructors center around the effect of the system on the teaching situation. A study done by Robert C. Birney, Professor of Psychology at Amherst College, on the "Effects of Grades on Students" reports that students think grades indicate too little about ability and potential; that they feel the need for extensive comments from their instructors instead of merely a letter grade. Students always consider their grades too low. They find the system too capricious and variable.1

These reactions are confirmed by students on the Berkeley campus. A set of interviews conducted with extremely bright students produced a list of what the students considered their grades to have measured besides academic achievement:

1. The interest of the student in the course.
2. Attendance.
3. Health.
4. Attitude.
5. Ability "to psych out the professor."
6. The outside activities of teaching assistants or professors.
7. Nerve, crassness, or aggressiveness: the aggressive student who goes to see his professor tends to do better at Berkeley.
8. The amount of accommodation a student is willing to make to satisfy another's wishes.
9. The degree of conformity or deviation from the professor's individual norm, academic or nonacademic.
10. The ability to write and communicate.
11. Nationality: the Hungarian student who writes English in a peculiar manner may be found either charming or irritating.
12. Religion: it may influence a student's ability to accommodate to the professor's assumptions. A section on Nietzsche with a "cool" teaching assistant finds the traditional believer back in the dark ages.
13. Sex.
14. Prep school background.
15. Parent's income: e.g., the ability to buy books and supplies; intellectual environment at home.
16. Dress and neatness.

The student interviewees objected to having a single letter grade stand for many things, with no indication of what the proportions were of neatness, punctuality, intelligence, effort, or achievement. They also noted that some of the indicated injustices in grading (advantages of prep school preparation, family environment and the like) tend to diminish with time, and hence they particularly recommended a grade-free beginning period of one or two years.

In one sense, the accuracy of the students' complaints about the arbitrariness of grading in each course is not important. It is not so important, surely, as the fact that students believe that the standards in most courses are not objective; they believe, by and large, that in order to get good grades it is necessary to cater to the specific prejudices of the individual professor; they complain that the grading system forces them into dishonesty both in intellectual and other ways; they claim that the regulative notion governing student conduct at the University is not the pursuit of knowledge, but playing the game. Though a student can remain here without becoming a "grade grubber," can he get into graduate school without it? Can he get graduate fellowships without it? They would be inclined to say that generally he cannot, due to lack of faculty contact that would provide a special recommendation for graduate school admission or fellowships.

These students believe that the active pursuit of grades, rather than learning, is essential and that much of it has nothing to do with intellectual achievement. In short, these students believe that adjusting their behavior to each professor in a Machiavellian way is the means to success, and they are doing it. The distortions in the teaching situation created by these students' cynical attitudes toward grades and their consequent alienation are incalculable.

There is also considerable evidence to indicate that such student reactions are not entirely their imaginations. Paul Dressel and Clarence H. Nelson, in Evaluation in Higher Education, write that a widely recognized problem with grading is the disagreement in actual fact about what a grade really measures:

> Absence, bad attitude, nonparticipation in class, and the like are considered by some teachers behavior contributing to a low grade. Grades can be and are used as a weapon to enforce temporary student conformity to prejudices of the teachers. These prejudices may have justification in the content and methods of the course, but they also may be purely personal and irrelevant. The case of an excellent student who was failed because his instructor disliked his attitude comes to mind. The inclusion of such personal, intangible, and often unspecified elements in grades makes their meaning unclear. Unfortunately, this very uncertainty causes students to focus on grades more than they otherwise would. We are all preoccupied by the uncertainties in our lives.[2]

Though Dressel and Nelson conclude that "Grading is inevitable," they go on to say "it is, at best, inaccurate and unreliable."[3]

Perhaps the most eloquent protest about the effects of teacher-assigned grades on the teaching situation comes from Joseph J. Schwab.[4]

> The work cannot, by the farthest stretch of the word, be called willing when it is done perforce under the whip of an imminent inquisition. It cannot be called independent when it must meet a test arbitrarily set by the same man who sets the work . . . It is fair to say that ingenuity could not combine the inimical effects of bread-and-butter love and submission to the taskmaster and inquisitor more effectively than does

the institution of the teacher-set examination. Without its removal, the possibility of establishing a sound teaching relation with the vast majority of students is well-nigh nil.

B. Grading - Accuracy and Uniformity

Osmond E. Palmer notes that "Studies have shown that an instructor does not usually give the same grade to a paper when he rereads it after an interval of time without knowing his previous grade," a potent argument for leaving only an analytic and evaluative comment on a paper and no grade.[5] If inaccuracies in the grading of students are indeed inevitable, they form a powerful argument against grading. Grades are the rewards and punishments of our institution. If they are dispensed in a fashion which is arbitrary in any significant degree, then the institution is not doing its job. "It is argued," writes Byron Stookey, Director of Academic Planning at Santa Cruz, "that grades are after all only a symbol, and that if the student is genuinely concerned with learning his attitude toward his education will be unaffected by grades. But this defense underestimates the impact of symbols. Symbols give meaning: grades give powerful meaning to a system of motivation derived from teachers, classmates, family, community, and our whole culture."[6]

Grades often are not as objective as teachers believe. There are hard and easy instructors. The widespread belief that grades are given on an absolute scale seems not to be valid; Lewis R. Aiken[7] has shown that college teachers usually grade with reference to the existing ability of their students; they "curve" their grades. In any given class most professors, consciously or unconsciously, have limited the number of A or D grades before they have met the class. Often the selection of students of higher ability is not followed by higher grades. Grading standards tend to shift with the ability level of the class, getting more stringent in each successive year.

A study of the relationship between personality and motivational factors and academic performance, reported by D. Black of the Counseling and Testing Center at Stanford, indicated several interesting trait patterns in men who do well academically. One of the most important might be called "cooperativeness," a trait including the tendency to be described as helpful, moderate, respectful, appreciative, sympathetic, and sensitive. Black notes that "this constellation of traits is much more characteristic of women than of men in our culture," and hence may account for men constituting seventy percent of the undergraduates at Stanford, but only 57 percent of the Dean's List. Men and women at Stanford, he claims, have the same general levels of ability and seem to spend the same amount of time studying. Black asks whether "women . . . are . . . beneficiaries of . . . unconscious bias on the part of the faculty?"

C. Grading and the Values of the University

Other complaints about the grading system have less to do with the teaching situation per se than with the educational situation as a whole. Mervin Freedman sees the American men's college encouraging a definition of success as "academic and later business or professional success; and achievement of this kind is becoming the sole standard by which middle-class boys and young men evaluate their worth and self-esteem."[8] He sees the grading system as a vital part of this process: encouraging this whole set of values, devaluing personal excellence and achievement apart from academic or vocational success. As enforcer of the work-success ethic, the grading system allows no time for creative leisure. Freedman notes that it is possible students "have absorbed this slave mentality and made

it a part of themselves. Students are uncomfortable, unless they are working hard or being overworked. There are those who would argue that the colleges are participating willy-nilly in the pressurized system of grading that begins with pre-school--trying to do well in order to eventually get into a good college and be 'successful.'" Those who criticize the grading system can well ask the questions, What are the long-term educational effects of our grading system? Are we not encouraging the conformity which is the constant complaint of American social critics? Are we not robbing the student of the leisure traditionally associated with the university in its European cradle? Freedman argues that what is needed in our society is "originality, not conformity."

Writing in The College Board Review, Paul Heist of the Center for the Study of Higher Education noted that "The development of the individual to the fullest extent possible is, of course, the most appropriate educational goal in a democracy."[9] In an interview with me he said that the present grading is too monolithic. It fails to encourage diversity by taking account of the valuable variety of qualities in our undergraduates. One of the things grades do, he noted, is encourage conformity, erecting one norm of academic performance and direction for all.

Serious tensions may develop in the teaching situation because of grades. A survey done by Paul Heist at a distinguished college found students, as at Berkeley, complaining of a distance between faculty members and themselves. And yet four freshmen in the top twelve percent of the class according to the grade point average complained that they felt unable to visit professors for fear of being criticized by their peers. It can be inferred that enough students cultivate faculty members to exploit personal contact for grades so that the more sincere students fear falling under the same accusation if they should get to know and get help from their professors. The students we consulted at Berkeley agreed with these inferences.

D. Grading and Motivation

One of our graduate students who is preparing an article on higher education has noted that "The grading system is based on a primitive psychology of the student; it assumes that he is best motivated by fear and the desire for status. But the drive for status as a means of overcoming inner fears and inner emptiness is at best a pitiful sham and at worst a cruel hoax; and the reliance on an abstract numerical or alphabetical designation to induce people to work ignores the simple fact that any human is most strongly motivated when someone whom he respects devotes attention and serious consideration to his work." John I. Goodlad, Professor and Director of the University Elementary School at UCLA confirms this student's speculation from laboratory results. "I take the position very firmly that the only meaningful reward in learning comes from learning itself and that all of these external rewards serve to contaminate the process."[10]

D. Black of the Counseling and Testing Center at Stanford reports that no definitive study has been made to say whether liberalized grading would reduce incentive, or increase it. Black thinks that since "sixty percent of all grade point averages fall within one grade point (2.0 and 3.0), many able students develop a fatalistic attitude which leads to performance below their capacities." Robert C. Birney reports that failing or near-failing grades produce great effort in the student, but that with higher grades, the effect is unpredictable. In a course of little interest, high grades lessen the amount of study. Birney tentatively concludes that high grades may induce pessimism and cynicism; the student who has attained high achievement feels that any increment in grade aspiration only

involves more effort, more politicking, more anxiety. Also, students often have a level of satisfaction: one may take himself as a B student, another as a C, and each will be content when that grade level is achieved.

As for the effect of low grades on bright students, Paul Heist in a report on grading problems prepared for another college indicates quite varied reactions. Some students tend to work harder, others do the opposite, and for some the effect of low grades is neutral. Speaking of the bright students who work less when punished by low grades, Heist says, "The reactions or 'mechanisms' resorted to are sundry and diverse, and for some these reactions lead to the first obvious steps toward a form of under-achievement or 'beating the system' (and themselves.) . . . An evaluative system should not worsen the learning environment, but serve the students' needs and development."

Joe L. Saupe, writing in Evaluation in Higher Education, writes that "Although the experimental evidence is slight, it appears safe to say that intrinsic motivation is to be preferred to extrinsic in promoting efficiency in learning. If the student's goal is to satisfy the teacher, he may do so, but he may learn course material only casually and incidentally. On the other hand, success in activities motivated intrinsically leads not only to desired learning outcomes but also to further engagement with these types of activities." Dressel and Nelson (in the same book) call attention to the disparity between the teacher's motivation and what he makes the motivation of the student: "Examinations and the grades derived from them should not entirely determine the amount and the kind of work that students do. . . . for the teacher, personal satisfaction and other even less tangible elements combine to motivate him to an effort above and beyond the call of the dollar. The student must have some of the same motivations, and it is one of the teacher's responsibilities to arouse them."

There is reason to believe that grade motivation not only produces very little in the way of real knowledge or discipline, it may be encouraging a kind of learning which has peculiar irrelevance in our time. Carl Rogers and other educators have noted that the kind of education appropriate until recently might have prominently featured the inculcation of stored knowledge already accumulated; what is needed now, he argues, is the inculcation of the ability to meet problems in a rapidly changing manner. Students must learn the process by which new knowledge is to be acquired--learning how to learn. In a constantly dynamic society, knowledge does not count for much, since every "subject" must be relearned in an ad hoc way all the time. Rogers argues (1) that knowledge is relatively useless because it dates fast; (2) students have an impressive rate of forgetting for knowledge acquired in relatively meaningless or hostile situations; (3) therefore the only way to inculcate the ability to learn the process of learning is to put students to work really learning. This is not necessarily uncontrolled learning, but it is much less highly regimented or externally imposed than what prevails at present. Rogers takes strong exception to the present method of encouraging education by offering extrinsic goals: "the student's desire to learn can be trusted . . . human beings have a natural potentiality for learning . . . Self-initiated learning, involving the whole person of the learner--feelings as well as intellect--is the most pervasive and lasting . . . Creativity in learning is best facilitated when self-criticism and self-evaluation are basic, and evaluation by others is of secondary importance."[11]

E. Grading and Graduate Admissions

There is significant evidence to show that the whole matter of graduate admissions is badly handled, and that the air of reliability that the grading system gives it protects

serious defects. One can point to the attrition rates between graduate school admissions and Ph.D.'s granted; U.S. Government estimates of the rate go as high as 20 to 1. These figures suggest that whatever system graduate admissions officers are using is not highly effective. In other words, though many graduate officers say that college grades predict success in graduate school and so should serve as an admission criterion, there is virtually no evidence for the assumption. It is just possible, particularly in the light of staggering attrition rates and the different kind of education that graduate schools are supposed to give, that we may be admitting the wrong people. One might consider the case of non-creative people who are high grade achievers; they come to the dissertation stage of the Ph.D. and they are unable to formulate a problem because their highly regimented education, dependent on outside stimuli, has not prepared them for a situation in which there is no assignment. This complaint was voiced by one of our professors of economics about many of his graduate students who no longer have a due date facing them. This crippled perspective on achievement can hinder a student later: he is often incapable right after college of motivating himself, or arranging his own time. T. R. McConnell, Director of the Center for the Study of Higher Education, testifies that grades do not work in measuring and predicting the kind of ability and performance that graduate schools are really interested in. And Edward Bowes, State-wide Director of Admissions, reports that there is very little evidence that <u>anything</u> used in determining graduate school admissions is successful.

By considering each student who presents himself for admission individually, another state university, the University of Michigan, is able to show dramatic reductions in the number of drop-outs among those who enter at the college level. Selection of the right candidates amply justifies the added cost.

While graduate schools tend to demand letter grades from colleges, they may be able to adopt to other systems. Howard M. Teaf, who organized a nationwide conference on grading two years ago, writes in a letter of August 23, 1965, that

> I am rather surprised that your inquiry has developed evidence of "pressures exerted by graduate schools for conventional methods of evaluation." At our conference, there was naturally much discussion of this point both among the colleges and with the representatives of graduate schools and of associations of professional graduate schools. We found that all the graduate schools are in favor of continuing use of grades and averages as reported by the undergraduate colleges as a basis for their administrative procedure and decision. But all of them said that they could "learn to live without grades." On the other hand the undergraduate college people were apprehensive about undertaking radical alteration of their rating systems <u>individually</u>. I think I am correctly reflecting their apprehension when I say that they did not fear intentional discrimination against a college that abandoned a formal grade system, but they did fear unintentional discrimination in favor of the student from the college where there was the more definitely known grade, given two students with otherwise similar records and from colleges roughly comparable in standing.

Should the Berkeley faculty decide to alter the grade system, there is considerable evidence (See Appendix, Responses to D letter) that other schools would do the same. If so, it is less likely that the graduate schools would discriminate against non-grading institutions. Definitive action by Berkeley might be decisive in encouraging the graduate schools to live without grades and to find more reliable criteria for admissions.

F. Grades and Transfers

Grades are also used in transfers, and more and more of our students are transferring. But Dr. Dorothy M. Knoell, who has been studying the transfer problem within the California system of higher education for some years, writes in a letter dated August 23, 1965, that even "a pass/fail system of grading need not handicap the applicant for transfer from the University, if discussions with the receiving institutions are planned at an early stage" in the process of switching from a conventional grading system. The problems in switching over are diminished at Berkeley by the possibilities for articulation that the state educational system allows. Dr. Knoell does not think that students transferring to prestige private institutions will have difficulty. Apparently colleges can learn to live without grades just as graduate schools can.

G. Grades and Employment

Grades are also used by businesses in hiring, but opponents of the grading system point out that their use by business is by no means universal; businesses can do without as many grades as we give them, and it is the complaint of many businesses that what they really want from faculty is not grades, but personal and individual evaluation of students. The last point suggests that every effort should be made to establish some system whereby businesses can get recommendations from faculty on individual students.

There is little real evidence of the correlation between grades and success in business. Demands for grades will differ from business to business; in certain areas, grades are scarcely used at all. A letter from Dean J. A. Zivnuska of the School of Forestry notes that "in regard to the professional employment of our graduates, grades are far less important than in the academic world . . . [employers] rely heavily on the recommendations of the faculty." Professor H. L. Vaughan of the College of Environmental Design writes "In the hiring of graduates of Departments of Architecture and Landscape Architecture, I think employers depend almost entirely upon exhibits of work rather than upon grades in courses. Our fields may be unique in that our students produce so much tangible evidence of ability or lack of it in the form of drawings and reports." Dean M. W. Morgan of the School of Optometry writes that employers seldom use grades in their hiring decisions, relying instead on personal recommendations about character, appearance, and the like. It is fair to say that employers seeking candidates to fill jobs closely related to academic work, like research, do use grades; other kinds of employers may or may not use them, but could probably do well without them.

Mrs. Nansi Corson, Acting Director of our Student and Alumni Placement Center, points out that the great difficulty employers have at Berkeley at present is that often the faculty either does not cooperate in filling out standard rating sheets on students that employers use, or does not know its students well enough to do so accurately. Two hundred and forty employers in a confidential survey of placement facilities on 57 college campuses in the West ranked Berkeley only average in its ability to furnish detailed faculty reports.

H. Grades and the Battle of Life

A complaint often heard is that the grading system encourages a kind of competition alien to the real purposes of university education. There is no reliable information about just how much competitiveness there is, and what its effects are. It is generally agreed, however, that there is a lot of it, and that it is not wholly a good thing.

Writing in the College Board Review, Mervin Freedman attributes part of the phenomenon of student alienation in colleges both large and small to the competitive situation the student is thrust into.

> Consider the situation of an entering freshman at a prominent college. He enters a complex social environment in which . . . he is unsure of himself and his abilities . . . Since he is in competition for grades with his fellow students, it is difficult to establish with them relationships in which he can relax and simply be himself. Of course, most students do recognize after a while that they can perform at least adequately within the system. . . . But even when this does occur . . . it is all too often a matter of anonymous socialization at the expense of intrinsic individual development.

Freedman yokes together "the atmosphere of competitiveness and isolation which has prevailed on our college campuses since the early 1950's."

The kind of competition prevailing in life does not resemble the one prevailing in school. The notion that grading prepares people for life is doubtful on other grounds as well. Even if grades were eventually able to predict success in different work roles, Junius P. Davis has noted, grades may never be an index of a student's potential for self-satisfaction in nonprofessional roles. At least there is no evidence to indicate that grading helps people lead the good life. Quite the contrary: it may be argued that the grading system discourages the development in the student of intrinsic and lasting intellectual interests and of self-definition generally.

Grades, says Mason Haire, encourage the student to believe that life is divided into neat portions separated by terminal situations. The feeling, which many of our students acquire, that "when the test is over the responsibility is over" is probably not a good preparation for life.

The continual presence of outside judges may cripple the student's capacity for self-evaluation, thereby encouraging conformity and general mindlessness later. The constant rhythm of reward in school may even contribute to an American's constantly collecting new status symbols. The American people is conditioned by the schools and specifically by grading and promotion to expect some sort of symbolic award or recognition regularly.

I. Grades and Creativity

Another complaint is that grades have a negative or low correlation with the trait of "creativity." One item of evidence supports the contention that grades are positively related to creativity: a study done by William K. Lebold of engineering graduates from Purdue showed a significant relation between the graduates who registered patents and college grades. However, much information points in the opposite direction. The study by Kelly previously cited on discrepancies between instructor grades and term-end exam grades in the same course showed that those students who got higher ratings from the instructor indicated on personality tests that they were more conforming, compulsive, rigid, and insecure; this would suggest that teachers bias their grades in favor of the conformists. Another study by Holland[12] shows that the nonintellectual factors most related to academic achievement are persistence, strong superego, and the like. Holland contrasts these with Cattell's list of characteristics for the creative person: intelligence, emotional maturity, dominance, adventurousness, sensitivity, introversion, radicalism, self-sufficiency,

tenseness, less subjection to group standards, impulsiveness and the like. The two sets of personality traits for achievers and for creative types are at odds: though two of the scales correlate, five are neutral, and five have a negative correlation with one another; the implication is that the college achiever has less potential for creativity than many nonachievers.

P. E. Jacob writes that "with few exceptions American students were about the same when they left college as when they entered." Attitudes did not change, values did not change, character and personality did not fundamentally change, nor were there significant intellectual changes. "The student may learn more about history or mathematics, but his intellectual approach to knowledge or learning is likely to remain conventional rather than innovative, orderly or systematic rather than creative."[13] Paul Heist has concluded that "the colleges usually do little to encourage creative expression and the full use of intellect on the part of those who do finish four years. Apparently none of the institutions we have been studying really succeeds either in understanding the promising, potentially creative individual or in providing an environment in which his gifts will flower." Heist says this lack is more apparent in the arts than the sciences, but concluded that both select liberal arts and science institutions "lose more creative youth than they educate." It is apparent that low grades, or the grading system in general, serves as one of the chief instrumentalities by which the college may indicate its lack of understanding of such promising students.

Evidence for low correlation between grades and creativity appears in virtually every field. In a study of creative achievement in both science and the arts, similar results are obtained. Creativity was defined here more or less as the winning of some public recognition for original work. Then the creative and noncreative students were given psychological tests, and various correlations were made. "Academic performance appears to be the function of . . . perseverance, self control, good behavior (good citizenship) . . . and rigidity . . . Creative performance seems to be the outcome of a conscious conception of being original, active participation in creative hobbies, an asocial temperament, independence, reinforcement by parents of similar values." Once again, academic achievement and creativity seem to be consequences of different qualities and dispositions.

College faculties appear to be well aware of this seeming discrepancy between grades and creativity. A study done by Junius A. Davis[14] at Amherst, Cal Tech, Cornell, Dartmouth, MIT, RPI, Rutgers, and Stanford showed that faculty evaluations of students on a scale of creativity had a low correlation with the grades that the same faculty members gave to the students. In this case it was not psychologists who were defining creativity, but the faculty themselves. A study done at the University of Michigan in the area of graduate psychology courses showed that undergraduate gpa had a .10 correlation with faculty prediction of professional contributions from the students, a .08 correlation with faculty predictions of contributions to science, and a .42 correlation with grades in graduate school.

J. Grading and Cheating

One of the effects of the present grading systems that critics deplore is cheating. "Several studies have suggested that . . . cheating . . . is widely prevalent in colleges."[15] It is hard to know if there is any final solution to this problem short of eliminating the present kind of evaluation. More personal evaluation, such as that at Bennington, Sarah Lawrence, and Goddard would eliminate it. A system where grading is

done on the basis of complex and sophisticated qualifying examinations in various areas would also help to control cheating.

K. Grading and Teaching

Grading gives teachers something to hide behind in dealing with their students. If you say the student failed a course because of reason X or for reason Y, you may have to justify the course, its procedures and so forth. If you say the student failed because he got a D or an F on a competitive examination, your statement is impersonal, relative to other students and to general though vague standards. It seems objective and not a condemnation of one human being. Grading protects bad teaching at present by allowing the teacher to demand of the student any effort whatever. Student criticism is silenced and the teacher is not forced except by the limitations of his conscience to think through what he is doing with his course. Max Marshall, as we have said, notes that grades influence teaching in a bad way because they herd students into following the teacher's path, and also because they influence teaching techniques and examinations. The questions on a test are too often meant to press parrot-response buttons rather than let the student think.[16]

II. PROPOSALS

The list of criticisms of the present system of evaluation is formidable and cannot be ignored by those interested in improving education at the University of California. It is clear that any change in the system which would not lead to serious disadvantages and which would eliminate some of the sources of criticism is highly desirable. Frequent suggestions for reform include increasing the number of distinctions, such as adding pluses and minuses, or going on a percentage system. An increase in the number of grading distinctions would answer virtually none of the objections critics of the present system have; including the objection of inaccuracy.

There are some promising reforms that the individual teacher can easily practice by himself. Dr. Ann M. Heiss of the Center for the Study of Higher Education told us in an interview that one fundamental improvement would be to have instructors minimize their reference to grades and testing. Her researches on education at Berkeley have shown that faculty are to a large extent responsible for the emphasis on grading that prevails here. Another reform that the individual teacher can practice, and that was uniformly suggested by our student consultants, is less grading and more diagnostic commentary on their work. They want criticism on bluebooks, papers, and other exercises--suggestions for improvement.

For possible reforms, there are two promising directions, and we should take both. We can develop more sophisticated, humane, and reliable systems of formal evaluation, that do full justice to individual achievement--as in the instructor's detailed written comment on the student's seminar performance. A collection of such comments, made in the course of a goodly number of small classes during the student's college career, would answer the problem substantially; it is a direction toward which we should move. Meanwhile, a simpler and readier improvement would be to reduce drastically the amount of systematic, formal grading we do, and to simplify our notation of it. We should undertake substantial experiments in pass-fail grading.

A. Pass-Fail Grading

Pass-Fail grading is no new thing at Berkeley. In 1941, honor students first began to have the option of taking three units in courses outside the major each semester of the junior and senior years. Henry Turdel, '41, a proponent of the plan said: "The pass or fail system will some day be offered to all students of the University . . . Under ordinary circumstances many students are afraid to go too far out of their major in selecting elective subjects . . . Through the pass or fail system, the student may take a course which he thinks will enlarge his general fund of knowledge without memorizing difficult and unessential details, and for which he need no longer suffer loss of grade points or grade point average."17 In 1942, both at Berkeley and elsewhere in the country, there was serious talk about pass-fail being extended to cover all courses so as to graduate more young men for the armed forces. All the arguments based on education were also raised for the pass-fail system, but nothing happened. Only now does it seem that Henry Turdel's prophecies are coming true. Cal Tech has been running a two-year experiment in which only pass-fail grades are given in the freshman year. Princeton is drawing up a list of courses which can be elected for pass-fail at the rate of one per semester. Harvey Mudd College will probably imitate the Cal Tech example and eventually put all four years on pass-fail if the experiment works. Cornell is drawing up plans for imitating the Princeton plan. Pomona already allows juniors and seniors to take a total of eight courses for pass-fail in nonmajor electives. Kenyon does the same.

Most worthy of our attention, two of our new campuses have taken the lead in this area of reform. Santa Cruz has opened this Fall with only pass-fail grades for 1965-1966. Irvine allows any course outside the major and related fields to be taken for pass-fail except specific courses required by departments or the university. This means that a student could take as much as half of his academic work under pass-fail, though it is expected that twenty percent will be more like it.

The present pass-fail option at Berkeley is statistically negligible in its effects. Last fall there was a total of only 983 pass grades given (out of a total of about 92,000 undergraduate grades) to 1325 students eligible to take a pass-fail course that semester.

Objections to pass-fail grading in general would be (1) it is still teacher-set grading. Grades have not been removed from the classroom; (2) the pass and fail grades will still suffer from inaccuracy and from variation with the instructor. However, the influence of special nonacademic biases will largely disappear when there are only two grades.

There are many anticipated improvements, on which a few comments are in order. We may expect, for example, that under a system in which the student has only to pass a course, that it will no longer be possible for the bad teacher to keep a captive audience under his control. There will be something resembling a free market for the better students, and we may expect that only those classes where real teaching is going on will be frequently attended. Under such a system, the student will have to learn to motivate himself; he will have to decide why he is in college; he will not be able to hide mindlessness under the mantle of a grade point average, but in the intellectual atmosphere of this college he will be forced to prove his intellectual worth to his peers in some meaningful way.

In the area of motivation such speculation is possible about what may happen. It is reasonable to expect that students will undergo a period of adjustment to the new system, in which there will perhaps be some loafing. But our better students are probably wise enough about their future lives and careers to learn how to value education for itself,

once the euphoria has worn off. Our poorer students will no more be able to loaf under pass-fail than they can at present.

In considering the use of the pass-fail system, the faculty has the choice of a number of alternatives.

1. Minor Enlargement of the Pass-Fail Option

We could choose to imitate schools like Princeton, Pomona, and others that are about to or have instituted the pass-fail option for a certain percentage of courses. The Princeton Plan will put a number of courses, usually of small enrollment, on a pass-fail basis each semester. These courses would be taken outside the major; they would allow for exploration in nonmajor fields. Though such a very cautious answer to the grading problem offers an opportunity not now available, it does not relieve the student of the burden of competition, anxiety, and the rest; nor does it change the teaching situation in general. It does not go very far in response to the faults of the grading system.

2. Pass-Fail in the Breadth Requirements

This improvement would eliminate grading in as much as a quarter of the student's work, thus carrying many of the benefits of a program for pass-fail in the first two years to be described. Furthermore, it would lead to valuable experimentation in the purposes of the breadth requirements. Under the present graded system, for a breadth requirement say, a student might know he could pass Chemistry and want to take it, but he might also know that he could get an A in Geology. It is reasonable to expect that under a uniform pass-fail system for breadth requirements he could feel freer to seek those courses which might really interest him and be in his judgment most valuable. The grade consideration would disappear from his sampling of courses outside his major field.

Another advantage of such a plan is that it might allow the professor some extra freedom. Without the need to grade students competitively, the professor would have less need to hold them to rigidly uniform assignments; he would be free to accept from students diverse modes of satisfactory achievement and thus might find more variety and interest for himself in this teaching of the course.

Grading only pass-fail in breadth requirements would allow the student to experiment outside his specialty all four years. It would also lend itself to interesting new educational directions. Great flexibility in arranging student schedules, and an atmosphere hospitable to a variety of field studies, would be created by the grading of general requirements in a simple, unified way.

3. Pass-Fail in all Nonmajor Courses

The advantages and disadvantages of this system may be calculated in large part by extrapolating them from the descriptions of previous systems. One notable objection: it would put a lot of pressure on the student for achievement in the major as measured by grades. This might be desirable if the faculty wishes to stress extrinsically disciplined performance in the major: under any pass-fail system with one or more courses outside the major, the major would be affected in fairly predictable ways. For example, students might "work harder," but they might lose independence and some self-motivation (both important factors in significant learning) in the major field which is graded on a coercive system as opposed to the grading system of other courses.

4. Pass-Fail in the First Two Years

This is the pass-fail proposal that seems most likely of being acceptable to faculty, to graduate schools, business, and others, and that would at the same time go far toward alleviating many of the difficulties of the present system. Mrs. Nansi Corson, at our Placement Office, assures us that there should be no difficulty under a pass-fail system for the first two years in dealing with business. Graduate admissions officers confess that pass-fail in the first two years will make virtually no difference in their consideration of applicants. Grades in the major and in the last two years are the decisive factor in graduate admissions.

In the same vein, in the interests of alleviating the adjustment problems of junior college students who come to Berkeley in the junior year, it might be useful to extend to them a pass-fail option for all courses in their first semester. Dr. Dorothy Knoell has pointed out that one-third of our JC transfers leave Berkeley before finishing, while only 13 percent are "required" to withdraw. She suggests that our JC transfers are severely burdened, often because advisors require the entering transfer student to take an extra heavy program with a view to determining his fitness.

5. Pass-Fail in all Four Years

This alternative allows individual exploration and achievement; relieves students of irrelevant pressure; puts the burden of educating himself on the student; and does away with tensions and distortions in the teaching situation. An extensive alternative system of evaluation might have to be invented (for example, qualifying examinations, or an expansion of the amount of recommendation writing) for certifying our students to graduate schools.

B. Comprehensive Examinations

One other important alternative the faculty should consider is the development of a system of qualifying examinations for majors and for breadth requirements. Such a system could operate with or without a pass-fail system for courses, and it could generate either conventional grades, or reports of student achievements in prose commentary, or both.

This method of examining was tried at the University of Chicago for many years. Some report that it was highly successful, but others who taught there or knew people who taught there report much complaint. It is possible that the basic Chicago model could be improved upon.

At the University of Chicago, examinations were written by a testing office composed of (1) members of the faculty in various disciplines (2) members of the faculty in the particular discipline being tested (3) psychologists, usually recruited from the faculty. With the help of pedagogical and psychometric expertise, examinations of a very high order of reliability and validity[18] were designed in a number of fields. Students received no grades in courses; all courses were optional; a student had to pass so many examinations in order to graduate and be certified as having completed the requirements of the University. Rather than making student and teacher antagonists in a game to see what grade the student might win, this system makes them comrades and colleagues in preparing for the examinations. Education was the goal. If the student complains about an examination under our present system, he can expect punishment; under the Chicago system, his complaint might be treated with more objectivity.

This system requires the faculty to really define, in advance, what the student is expected to learn. Chicago made its instructors define the passing, failing, and honors levels of performance before the tests were administered. The student then was graded on his mastery of skills and subject matter.

The system allowed the college to define certain broad unified goals for education, as well as encouraging the various departments to do so in more detail. It was required that there should be certain notions common to writing all the examinations, e.g., (a) the student understands the basic problems and approaches of the discipline, (b) the student has basic information. In preparing for the examinations, the student was given various resources: courses he might take, syllabi, reading lists, and the option to take practice examinations which were analyzed but not graded. The student could then check on his rate of progress. Examinations were highly complex and flexible, and could include four or five hundred "objective questions," open book essays, library assignments in research, laboratory assignments, and so forth. At the end of his college career a student would have one grade in the major, another in the minor and six in general education (humanities, English composition, social science, biological science, physical science, mathematics) and then a number of pass-fail grades in the remaining electives.[19]

Such a system does not do away with grading pressure on students; it replaces grading trauma with one great final exam trauma. To meet this objection partly, we might offer the examinations in parts, with provision for "second chance" reexaminations.

A comprehensive examination system would be a particularly logical supplement to a pass-fail system in the courses for the major. Irvine and Santa Cruz are both developing comprehensive examination systems, and Irvine is working with the Educational Testing Service in this area. It has been suggested that the resources of the State-wide University give adequate manpower for setting up a system of outside examiners for a given campus. This does not mean that the examination in the physics major, say, would have to be uniform throughout the state; only that a given campus could draw on examiners from other campuses to help it run its examination system. We are lucky in having such a resource, and it may be useful in our assessment of the validity of our specific educational offerings.

FOOTNOTES

[1] "College Grading Systems: A Journal Symposium" in Journal of Higher Education, February, 1964.

[2] The Committee on the Teaching of Botany in American Colleges and Universities of the Botanical Society of America, writing in An Exploratory Study of the Teaching of Botany in the Colleges and Universities of the United States (1938), p. 33 noted: "Pleasing personal behavior, self-assurance, apparent interest and enthusiasm on the part of the student may sometimes be mistaken for achievement in the course; unattractiveness, timidity, and an apparent indifference for the lack of it. Objective impersonal evidence of student achievement is needed." This is particularly interesting in the light of claims by those teaching in the sciences to a greater degree of objectivity in their grading. It is possible that such objectivity exists, but apparently only relatively. Testimony that leads one to the same conclusion comes from a paper written by William G. Perry, Jr., for the Bureau of Study Counsel, Harvard University, "Examsmanship and the Liberal Arts: an Epistemological Inquiry." Perry reports that at least one student of "considerable talent" gave him an "impressive analysis of the art of amassing 'partial credits' on examinations in advanced physics. His presentation confirms the impression that instructors of Physics frequently honor on examinations operations structurally similar to those requisite in a good essay." Perry's conclusion is that it is possible to "bull" as he puts it, in advanced Physics as well as in English. One might argue that in advanced Physics grading an examination may not be the open and shut matter it is often thought to be, and as in English, what one professor takes as ingenuity on an essay another may take as stupidity.

[3] Paul Dressel and Associates, Evaluation in Higher Education, p. 251.

[4] "Eros and Education: A Discussion of One Aspect of Discussion," Journal of General Education, VIII (October 1954), 51-71.

[5] Dressel and Associates, op. cit., p. 208. In a letter to the Committee, William K. LeBold, Assistant to the Dean of Engineering at Purdue, writes: "I think it can be demonstrated that grades in individual courses are often rather unreliable."

[6] Letter to the Committee, June 17, 1965.

[7] Educational and Psychological Measurement, 23 (1963), pp. 319-22.

[8] "The Men's Colleges and the Image of Man" in Journal of the American Association of University Women, 57 (1964), 107-110.

[9] "Personality Growth in the College Years" in The College Board Review, Spring, 1965, 56, 25-32.

[10] Letter to the Committee dated August 4, 1965.

[11] "The Facilitation of Significant Learning," to be published as a chapter in Contemporary Theories of Instruction, ed. Lawrence Siegel, by the Chandler Publishing Company. Rogers is now at the Western Behavioral Science Institute at La Jolla. Harold Taylor also notes in his letter to the Committee of August 2, that "discussion and independent inquiry are the natural and most effective ways to learn. They do not require the continual presence of the teacher or any kind of educational authority to be effective."

[12] J. F. Holland, "Prediction of College Grades from Personality and Aptitude Variables," Journal of Educational Psychology, 51 (1960), pp. 245-54.

[13] P. E. Jacob, Changing Values in College, New Haven, 1956, cited in a paper by T. R. McConnell, "Student Selection and Development in American Higher Education."

[14] "Summary of Major Findings from the College Student Characteristics Study," Educational Testing Service, Princeton, 1964.

[15] Paul Dressel and Clarence H. Nelson, Evaluation in Higher Education, p. 237.

[16] Max S. Marshall, "To Grade or Not to Grade," Journal of Higher Education, 23 (1952), pp. 264-269.

[17] The Daily Californian, January 17, 1941.

[18] Reliability and validity are two concepts psychologists have developed in the area of testing. A test is reliable when it works the same on two different occasions; a test is valid when you are measuring what you think you are measuring.

[19] My information about Chicago comes from conversations with several people, including Ralph Tyler who headed the examination office there for some years, an article by B. S. Bloom, "Changing Conceptions of Examining at the University of Chicago," in Evaluation in General Education, ed. Paul L. Dressel, and recent correspondence with the University.